Genome Sequence of "*Candidatus* Microthrix parvicella" Bio17-1, a Long-Chain-Fatty-Acid-Accumulating Filamentous Actinobacterium from a Biological Wastewater Treatment Plant

Emilie E. L. Muller,[a] Nicolás Pinel,[a,b] John D. Gillece,[c] James M. Schupp,[c] Lance B. Price,[c] David M. Engelthaler,[c] Caterina Levantesi,[d] Valter Tandoi,[d] Khai Luong,[e] Nitin S. Baliga,[b] Jonas Korlach,[e] Paul S. Keim,[c,f] and Paul Wilmes[a]

Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg[a]; Institute for Systems Biology, Seattle, Washington, USA[b]; TGen North, Flagstaff, Arizona, USA[c]; Water Research Institute, CNR, Monterotondo, Italy[d]; Pacific Biosciences, Menlo Park, California, USA[e]; and Northern Arizona University, Flagstaff, Arizona, USA[f]

"*Candidatus* Microthrix" bacteria are deeply branching filamentous actinobacteria which occur at the water-air interface of biological wastewater treatment plants, where they are often responsible for foaming and bulking. Here, we report the first draft genome sequence of a strain from this genus: "*Candidatus* Microthrix parvicella" strain Bio17-1.

Actinobacteria of the genus "*Candidatus* Microthrix" are commonly found in biological wastewater treatment plants, where they are considered notorious in causing solid-liquid separation problems and foaming (see reference 14 for a recent review). On culture media they grow meagerly, yet they seem to have an *in situ* competitive advantage mediated by their avid uptake of long-chain fatty acids, which are accumulated as neutral lipids under anaerobic conditions and converted into phospholipids for cell division under aerobic conditions (1, 13).

Here, we present the draft genome sequence of "*Candidatus* Microthrix parvicella" Bio17-1, a strain isolated from a Dutch wastewater treatment plant serving fish industries (10). First, two sequencing libraries were prepared from genomic DNA with mean insert lengths of 350 bp (paired ends) or 2,750 bp (mated pairs) and sequenced on an Illumina Genome Analyzer II. Raw 100-bp reads were error corrected with Quake (8). A total of 5.84×10^6 paired-end and 1.12×10^6 single-end reads with a minimum mean quality value of 30 and a minimum length of 70 bp were used for assemblies. Second, 24,031 single molecule, real-time (SMRT) sequence reads were obtained on a Pacific Biosciences PacBio *RS* using C1 chemistry. Error correction yielded 2,625 reads (232 to 1,984 bp).

Using the Illumina sequence reads, two preliminary assemblies were obtained with Velvet (17) and Edena (7) and merged with the minimus2 utility (16). The resulting 27 contigs were scaffolded with SSPACE (3), and gaps were filled with GapFiller (4). Additional assemblies were obtained using SOAPdenovo (11) (kmer values between 65 and 81, steps of 2) and CABOG (12). Error-corrected PacBio reads (9) were mapped onto the preliminary assemblies. Draft contigs were broken where discrepancies among assemblies or PacBio reads suggested misassemblies. Conversely, contigs were joined where contig ends overlapped with perfect identity for at least 500 bp. Manual curation of the assemblies was performed using Consed (5). Automatic annotation and draft metabolic reconstruction were performed by the RAST server (2). CRISPR loci were identified using CRISPRFinder (6).

The draft assembly consists of 4,202,850 bp, arranged in 13/16 scaffolds/contigs, with a mean GC content of 66.4%. Automated annotation identified 4,063 coding sequences, in addition to 1 rRNA operon and 46 tRNAs covering all amino acids. A complete pentose phosphate pathway and tricarboxylic acid (TCA) cycle are encoded in the genome. As previously hypothesized for "*Candidatus* Microthrix parvicella" strain RN1 (15), a nitrate reductase is encoded by the genome, but no nitrite reductase appears to be present. The strain is also predicted to be a prototroph for all amino acids, to be able to polymerize/depolymerize polyhydroxybutyrate, to accumulate polyphosphate, and to translate several selenoproteins. No genes are annotated that are related to photosynthesis. The assembly contains one CRISPR locus with 88 spacers.

"*Candidatus* Microthrix parvicella" Bio17-1's ability to process and accumulate excessive amounts of fatty acids is highlighted by its gene content: the genome encodes 28 homologs of long-chain fatty acid–acyl coenzyme A (acyl-CoA) ligase and 17 of enoyl-CoA hydratase. The genetic inventory of "*Candidatus* Microthrix parvicella" makes it of particular interest for future wastewater treatment strategies based around the comprehensive reclamation of nutrients and chemical energy-rich biomolecules.

ACKNOWLEDGMENTS

REFERENCES

1. **Andreasen K, Nielsen PH.** 1998. *In situ* characterization of substrate uptake by Microthrix parvicella using microautoradiography. Water Sci. Technol. **37:**19–26.
2. **Aziz RK, et al.** 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics **9:**75.
3. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27:**578–579.
4. **Boetzer M, Pirovano W.** 2012. Toward almost closed genomes with GapFiller. Genome Biol. **13:**R56.
5. **Gordon D, Abajian C, Green P.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.
6. **Grissa I, Vergnaud G.** 2007. CRISPRFinder: a Web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. **35:**W52–7.
7. **Hernandez D, François P, Farinelli L, Østerås M, Schrenzel J.** 2008. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. Genome Res. **18:**802–809.
8. **Kelley DR, Schatz MC, Salzberg SL.** 2010. Quake: quality-aware detection and correction of sequencing errors. Genome Biol. **11:**R116.
9. **Koren S, et al.** 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat. Biotechnol. **30:**693–700.
10. **Levantesi C, et al.** 2006. Phylogeny, physiology and distribution of "*Candidatus* Microthrix calida," a new *Microthrix* species isolated from industrial activated sludge wastewater treatment plants. Environ. Microbiol. **8:**1552–1563.
11. **Li R, et al.** 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. **20:**265–272.
12. **Miller JR, et al.** 2008. Aggressive assembly of pyrosequencing reads with mates. Bioinformatics **24:**2818–2824.
13. **Nielsen PH, Roslev P, Dueholm TE, Nielsen JL.** 2002. Microthrix parvicella, a specialized lipid consumer in anaerobic-aerobic activated sludge plants. Water Sci. Technol. **46:**73–80.
14. **Rossetti S, Tomei MC, Nielsen PH, Tandoi V.** 2005. "Microthrix parvicella," a filamentous bacterium causing bulking and foaming in activated sludge systems: a review of current knowledge. FEMS Microbiol. Rev. **29:**49–64.
15. **Tandoi V, Rossetti S, Blackall L, Majone M.** 1998. Some physiological properties of an Italian isolate of Microthrix parvicella. Water Sci. Technol. **37:**1–8.
16. **Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M.** 2011. Next generation sequence assembly with AMOS. Curr. Protoc. Bioinformatics **33:**11.8.1–11.8.18.
17. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18:**821–829.